VAA-Weekly-Progress

4/22-4/29

Context

- Worked on report draft

Goals

- Finish Introduction - Claire
- Related Work
    - AP-10k - Medha
    - Animal Pose - Zian
    - OpenApePose - Josh
    - 300 faces - Claire
- Finish Biological Definitions - Medha
- Embed Biological Definitions - Medha
- Adding taxonomical similarity work(from last semester) to the report - Claire
- Add point about how side-pose for species similarity does not help - Claire
- Do a Biological Experiment (downsized to 17 keypoints) - Medha
- Do a Biological Experiment (21 keypoints) - Parth
    - If it works, do Visible as well
- Description of Centroid Variation on Chimps - Parth
- Use OKS on all experiments to find which keypoints are better - Zian
- Testing species similarity at different k-values - Shaan
- Writing Orb into the Final Report - Shaan
- Dino with Stable Diffusion - Josh

Visual Representation of Biological Keypoints

Red - AP10k

Blue - Biological

Bio set was downsized from 21 to 17 keypoints, by removing the bio knee and bio elbow, and renaming the bio ankle to the knee, and bio wrist to the elbow

All 3 Definitions visualized

Red - AP10k

Blue - Biological

Green - Visible

Visible vs Bio vs AP10k Comparison

- Bio set was downsized from 21 to 17 keypoints, by removing the knee and elbow, and renaming the ankle to the knee, and wrist to the elbow

Metric	Visible	Bio	AP10k
coco/AP	0.725	0.622	0.692
coco/AP .5	1.000	0.935	0.935
coco/AP .75	0.842	0.726	0.782
coco/AP (M)	-1.000	-1.000	-1.000
coco/AP (L)	0.725	0.622	0.692

Centroid Metric: Mouse Experiment

- We wanted to try the centroid metric with smaller animals (so I picked mouse)

- The procedure was as follows:
    - Separated mouses from AP-10k into a testing set
    - Redistributed AP-10K without mouse with train/val/test -> train/val
    - Found top 10 species with centroid metric for mouses and created dataset
    - Downsized AP-10k to the same size as top 10 species set (1973) images

Model Testing Results

AP - Average Precision, AR - Average Recall on all mouse images

	Full AP-10k	Downsized AP-10k	Top 10 Centroid
coco/AP:	0.624	0.453	0.536
coco/AP .5:	0.917	0.860	0.889
coco/AP .75	0.715	0.451	0.559
coco/AP (M):	0.663	0.315	0.503
coco/AP (L):	0.624	0.458	0.537

	Full AP-10k	Downsized AP-10k	Top 10 Centroid
coco/AR:	0.677	0.526	0.591
coco/AR .5:	0.928	0.880	0.909
coco/AR .75	0.766	0.555	0.641
coco/AR (M):	0.700	0.317	0.517
coco/AR (L):	0.676	0.533	0.594

DINOv2+CLIP vs. DINOv2+SD

I was accidentally using the CLIP model instead of Stable Diffusion(For some reason thought that the CLIP was just the feature extraction backbone of SD)

Implemented DINOv2+SD processing of images with masks around each keypoint and with with a mask just around the animal:

Lists(Buffalo and Bison ranked low in both):

Whole Body Mask(COS) = ["squirrel", "rabbit", "mouse", "moose", "skunk", "dog", "raccoon", "sheep", "otter", "deer"] (Spider Monkey and Monkey ranked high, but not Gorilla)
Key Point Based Mask(KNN) = ["deer", "fox", "dog", "moose", "weasel", "mouse", "rat", "cow", "cat", "bobcat"]

Whole Body Mask(COS):   AP: 0.768   AR: 0.794

Key Point Based Mask(KNN):   AP: 0.775   AR: 0.798

Conclusion:

- "the task of keypoint estimation, unlike semantic correspondence, relies heavily on understanding what each part of an animal is/identifying a keypoint location given a certain image patch of the animal vs. understanding how each keypoint relates spatially to each other. This fact is further supported by the only successful experiment to lower the AR and AP results was the addition of the keypoint skeletons, which again focused on adding more spatial keypoint relation data instead of constricting the amount of data the model receives to only what is needed to place each keypoint separately."\
- Applicable to explain better performance of centroid variance(more focused on individual keypoints) vs. limb ratios(more focused on relations between keypoints)

Object keypoint Similarity

$$OKS = \frac{\sum_{i \in [0, N-1]} exp(\frac{-d_i^2}{2s^2\kappa_i^2})\delta(v_i > 0)}{\sum_{i \in [0, N-1]} \delta(v_i > 0)}$$

$- s : Object\ scale$

$- d_i : Distance\ of\ predicted\ joint\ i\ from\ ground\ truth$

$- \kappa_i : Per\ keypoint\ constant\ that\ controls\ falloff$

$- v_i : Visibility\ flag.$

Metric (%)	Full AP10k	Downsized AP10k	Centroid Variation	DINOv2+CLIP	Limb Ratios	Human Ranking	Taxonomic Species
Avg. Precision on Antelope	81.8	75.2	82.5	80.4	79.4	82.1	77.4
Object Keypoint Similarity	55.5	53.9	56.6	55.9	56.0	55.9	54.7

Personal Progress

Parth

- Worked on mouse experiment with centroid metric
- Wrote up the centroid experiments (chimpanzee and mouse)
- Worked on finishing VIP items

Medha

- Worked on 17 keypoint biological experiment
- In paper worked on biological keypoints, AP10k related work, biological definition embedding
- Worked on PD activities